БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. Как технологии распознают наши эмоции и почему это так перспективно [Электронный ресурс]. – Режим доступа: <https://rb.ru/opinion/tehnologii-i-emptsii/>. – Дата доступа: 20.05.2019.
2. About facereader [Electronic resource]. – Mode of access: <https://www.noldus.com/facereader/measure-your-emotions>. – Date of access: 20.05.2019.
3. Deep Learning [Electronic resource]. – Mode of access: - <https://www.affectiva.com/how/deep-learning-at-affectiva/>. – Date of access: 20.05.2019.
4. Dataset fer2013 [Electronic resource]. - <https://www.kaggle.com/deadskull7/fer2013>. – Date of access: 20.10.2019.

ИСПОЛЬЗОВАНИЕ ГРАФОВОЙ СХЕМЫ ДЛЯ ОРГАНИЗАЦИИ ХРАНЕНИЯ И ОБРАБОТКИ ДАННЫХ МЕДИЦИНСКИХ СИСТЕМ

А. В. Курочкин, В. С. Садов

Белорусский государственный университет, Минск, Беларусь

E-mail: alex.v.kurochkin@gmail.com

В работе рассматривается возможность применения графовой модели для организации централизованного хранилища по агрегации информации из различных медицинских информационных систем. Показано, что такая агрегация позволяет единообразно обрабатывать информацию о проведенных медицинских исследованиях как с точки зрения истории болезни отдельного пациента, так и с точки зрения статистической обработки исторических данных в рамках конкретного обследования. Также рассмотрены преимущества графовой схемы для данной задачи по сравнению с традиционными реляционными СУБД.

Ключевые слова: медицинские информационные системы; графовые базы данных; нереляционные базы данных.

Графовые модели в системах долговременного хранения информации представляют удобный механизм организации и обработки различных видов данных. Простота организации, отсутствие привязки к строгой схеме данных и возможность расширения без дополнительных миграций делают графовые модели оптимальным решением для работы со слабоструктурированными и денормализованными данными из различных обособленных источников. В данной работе рассматривается возможность применения графовой модели для моделирования схемы данных в рамках медицинской системы учета пациентов с необходимостью дополнительной статистической обработки информации в рамках отдельных исследований с использованием специализированных баз знаний.

Одной из ключевых особенностей современных медицинских систем является крайне слабая интеграция и отсутствие единой организации потоков данных в различных диагностических, терапевтических и административно-учетных информационных системах в рамках учреждения здравоохранения [1]. Например, большинство современных систем томографического обследования поддерживает экспорт информации только в формате DICOM, после чего результаты, как правило, записываются на внешний носитель, например, компакт-диск. При этом дополнительная информация о комплементарном медицинском диагностическом обследовании может храниться в рамках другой информационной системы, связанной с этим конкретным обследованием, а остальная история болезни – в отдельной централизованной базе по учету пациентов.

Во многих случаях замена информационных составляющих ради унификации обработки информации не представляется возможной, т. к. диагностическое оборудование может не иметь возможности подключения к компьютеру или к централизованной сети, либо может полагаться для работы с данными на специализированное программное обеспечение. С другой стороны, важной задачей является агрегация данных из различных источников в едином виде для упрощения работы как в разрезе истории болезни отдельного пациента, так и с точки зрения исторической статистической информации о каких-либо обследованиях или конкретных патологиях. Следует отметить, что введение агрегированного хранилища не ставит перед собой цель замены хранилищ существующих систем; вместо этого, оно может использоваться как дополнительный индекс для эффективной поддержки более сложных выборок и, в простейшем случае, может быть доступным пользователю только для чтения.

Графовые модели в рамках систем управления базами данных подразумевают хранение информации о сущностях, атрибутах и связях в виде вершин и рёбер некоторого ориентированного мультиграфа. Как правило, вершинам соответствуют некоторые объекты, а при помощи рёбер описываются отношения, в которых эти объекты состоят. При этом, в отличие от традиционных схем данных, сущности, ассоциирующиеся с вершинами, и сущности, ассоциирующиеся с рёбрами, являются равноправными [2].

Централизованная графовая схема для агрегации данных из различных медицинских систем может использоваться с точки зрения доступа к информации о пациентах и об отдельных обследованиях. Основное преимущество такой модели состоит в возможности единообразного извлечения информации с точки зрения конкретного обследования или с точки зрения конкретного пациента с одинаковой вычислительной

сложностью. В рамках графовой схемы каждый пациент и каждый вид обследования или диагностического оборудования могут быть представлены вершинами; в этом случае рёбра между конкретным пациентом и конкретным видом исследования могут быть интерпретированы как протокол прохождения этим пациентом данного обследования.

Поскольку графовая схема данных основана на мультиграфе, между двумя вершинами может находиться произвольное количество рёбер, то есть все узлы всегда находятся друг с другом в отношении «многое ко многому». При необходимости получения полной истории болезни и результатов всех обследований в качестве точки входа для запроса может использоваться узел, связанный с конкретным пациентом, а вся соответствующая информация может быть получена из рёбер, связывающих пациента с интересующими видами обследований; при этом требуется обход графовой модели лишь на единичную глубину. Такая агрегация наиболее информативна с точки зрения специалиста, проводящего комплексную диагностику. С другой стороны, при использовании в качестве точки входа некоторого обследования по конкретной патологии, обходом графа на единичную глубину может быть получена информация о всех исторических данных, связанных с этим обследованием. Такая информация может использоваться для статистического анализа, выявления общих исторических закономерностей, проверки статистических гипотез, а также для построения экспертных систем и систем поддержки принятия решений и для подтверждения корректности работы таких систем.

Основное преимущество графовой схемы данных по сравнению с традиционной реляционной является поддержка произвольной схемы данных в рамках атрибутов узлов или связей, т.к. схема данных для конкретного обследования определяется используемым форматом хранения и является уникальной для каждого вида обследования; кроме того, вычислительная сложность обхода графа относительно найденного узла на единичную глубину намного ниже сложности операции объединения неизвестного числа таблиц, что позволяет обеспечить большую производительность таких систем по сравнению с реляционными СУБД [2].

По виду хранимой и передаваемой информации медицинские информационные системы можно условно разделить на полностью аппаратные системы с экспортом на внешний носитель, аппаратно-программные системы с некоторым форматом хранения, программные системы хранения исторической и статистической информации и централизованные системы учета информации о пациентах. Для каждого из таких источников выгрузка информации для дальнейшей агрегации осуществляется исходя из особенностей конкретной системы. Для чисто аппаратных си-

стем выгрузка осуществляется вручную, посредством носителей, поддерживаемых конкретным оборудованием, после чего производится разбор целевого формата и агрегация полученных данных в центральную схему. Для программно-аппаратных систем и программных систем хранения выгрузка может иницироваться автоматизировано (например, по некоторому расписанию), а также по факту модификации встроенного программного хранилища системы. Наконец, централизованные системы могут быть синхронизированы в соответствии с возможностями об асинхронном оповещении об изменениях.

Организация извлечения данных из обособленных медицинских систем возможна только при наличии соответствующих программных интерфейсов или при использовании открытого формата, позволяющего получить информацию из конкретной системы в нужном виде. К наиболее распространённым форматам хранения можно отнести формат DICOM для работы с медицинскими изображениями, различные схемы разметки XML (в рамках стандартов ASC X12, HL7, CCR, openEHR и др.), а также табличные форматы, которые заполняются специалистами вручную (CSV, Microsoft Excel, файлы баз данных Microsoft Access) [3]. Во многих случаях одна и та же информация может дублироваться в хранилищах различных систем, не иметь строгой схемы и нормализации. При этом для объединения процесса экспорта данных в централизованную модель могут использоваться различные стратегии агрегации, например, механизм формальных проекций [4].

Таким образом, предложенный подход к организации агрегированного хранилища для различных медицинских информационных систем позволяет единообразно осуществлять выборку релевантной информации относительно истории болезни конкретного пациента и относительно исторических данных по конкретному обследованию, является намного более гибким из-за отсутствия требования скалярности и строгой типизации атрибутов, не требует миграций при изменении схемы, используемой в рамках конкретного обследования, а также обеспечивает большую вычислительную эффективность для типовых запросов, по сравнению с традиционной реляционной схемой хранения данных. Кроме того, извлечение и агрегация данных с отдельных систем может быть организована прозрачным образом, т. е. предложенное централизованное хранилище на основе графовой схемы подразумевается для использования в совокупности с уже существующими системами, а не в качестве их замены. При этом важной задачей остаётся приведение данных из различных источников к общему виду.

Предложенный подход рассматривался в рамках интеграции экспертной системы по диагностике многоплодных беременностей на базе УЗ

РНПЦ «Мать и дитя» [5], а также экспертной системе по прогнозированию и диагностике дегенеративных оптиконеуропатий на базе отделения офтальмологии 3-й городской клинической больницы им. Е.В. Клумова [6].

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. Горюнова В. В., Горюнова Т. И., Кухтевич И. И. Современные тенденции в развитии медицинских информационных систем // *Фундаментальные исследования*. 2015. №5, ч.1. С. 58–62.
2. Robinson I., Webber J., Eifrem E. *Graph Databases: New Opportunities for Connected Data* // O'Reilly Media, 2nd edition. 2015. 238 p.
3. Park Y., Shankar M., Park B.-H. et al. Graph databases for large-scale healthcare systems: A framework for efficient data management and data services. // *IEEE 30th International Conference on Data Engineering Workshops (ICDEW)*. 2014. P. 12–19.
4. Чернявский А. Ф., Курчокин А. В., Садов В. С. Обработка и индексирование денормализованных и слабоструктурированных данных // *Прикладные проблемы оптики, информатики, радиофизики и физики конденсированного состояния: материалы пятой Междунар. науч.-практ. конф.*, Минск. 2019. С. 164–165.
5. Прибушеня О. В., Курочкин А. В. Оценка плацентации при многоплодной беременности с использованием современных экспертных компьютерных программ // *Современные перинатальные медицинские технологии в решении проблем демографической безопасности*. Минск. 2017. 10 вып. С. 106–111.
6. Качан Т. В., Курочкин А. В., Головатая Е. А. и др. Роль искусственных нейронных сетей в выявлении ранней гибели ганглионарных клеток сетчатки у пациентов с дегенеративными оптиконеуропатиями // *Офтальмология. Восточная Европа*. Минск. 2019. Т. 9, № 4. С. 445–458.

ПРИМЕНЕНИЕ ЦИФРОВОЙ ОБРАБОТКИ ИЗОБРАЖЕНИЙ ДЛЯ РЕШЕНИЯ ЗАДАЧ БИОМЕТРИЧЕСКОЙ ИДЕНТИФИКАЦИИ ПО РИСУНКУ ВЕН ЛАДОНИ

В. А. Мельников, К. С. Шпак

*Белорусский государственный университет, Минск, Беларусь
E-mail: valik-m@list.ru*

Рассмотрен процесс обработки рисунка вен, включающий два этапа: предобработка входного изображения, локализация вен. Показано что для бинаризации изображения рисунка вен лучше подходят алгоритмы адаптивной бинаризации. Реализованы алгоритмы цифровой обработки изображений необходимые для локализации рисунка вен на изображении ладони.

Ключевые слова: цифровая обработка изображений; биометрическая идентификация; бинаризация изображения; детектор границ Канни; утоньшение линий.